

CrossMark  
click for updatesCite this: *RSC Adv.*, 2015, 5, 23922

# Structural investigations by *in silico* modeling for designing NR2B subunit selective NMDA receptor antagonists

Vishal P. Zambre,\* Varsha A. Hambarde, Nilesh N. Petkar, Chinmay N. Patel and Sanjay D. Sawant

Glutamate *N*-methyl-*D*-aspartate (NMDA) receptors are widely distributed throughout the mammalian central nervous system (CNS). They play key roles in many brain disorders, such as brain trauma, seizures, pain, Parkinson's and Huntington's disease. NMDA receptors are co-assemblies of four subunits. The selectivity of ligands for the NR2B subunit, out of the others, has shown the best side effect profile. The objective of the current studies is to establish the structural requirements for pyrazine and related derivatives for being NR2B selective NMDA receptor antagonists. 2D- and 3D-QSAR models have been developed. 2D-QSAR analyses revealed the key role of Baumann's alignment independent topological descriptors, such as T<sub>2\_F\_7</sub>, T<sub>C\_O\_7</sub> and T<sub>T\_T\_6</sub>, in biological activity prediction. Furthermore, 3D-QSAR analyses showed that steric and electrostatic molecular descriptors have significant influence in the optimization of pyrazine and related lead compounds as NR2B selective NMDA receptor antagonists. 3D points generated using a 3D-QSAR study have been studied to uncover minimum structural requirements of pyrazine derivatives. The significantly high prediction power of the best QSAR model supports the mechanism of NMDA receptor antagonism through NR2B subunit selectivity. A molecular docking study of the most active compound **18** revealed important structural insights needed to optimize pyrazine derivatives for use as NR2B selective NMDA receptor antagonists. Moreover, a putative structure-based pharmacophore model was established, having features of two aromatic rings, one hydrogen bond donor and one hydrogen bond acceptor. This pharmacophore model could be used as a query for virtual screening of commercially available chemical databases to identify new hits, and 2D- and 3D-QSAR models could be used to predict the activities of identified hits.

Received 19th January 2015  
Accepted 18th February 2015

DOI: 10.1039/c5ra01098e

[www.rsc.org/advances](http://www.rsc.org/advances)

## 1 Introduction

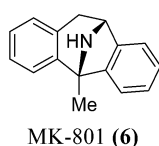
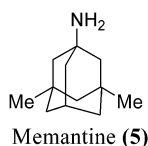
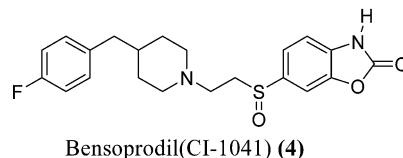
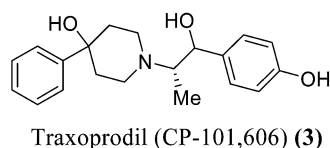
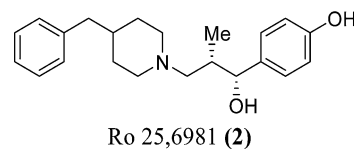
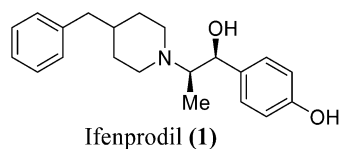
*N*-Methyl *D*-aspartate (NMDA) plays an important role in physiological processes, such as neuronal development, learning and memory, motor activity, and nociception, and also in pathological states of the CNS including strokes, seizures and pain.<sup>1,2</sup> NMDA receptors (NMDARs) are of great pharmacological interest and implicated in various disorders, such as Alzheimer's, Huntington's, and Parkinson's diseases, and also in neuropsychiatric illnesses such as disorders induced by alcohol and psychotropic drugs.<sup>3</sup> NMDARs are heteromeric assemblies of subunits consisting of NR1, NR2 and NR3 genes.<sup>4</sup> The NR2 subunit family is further divided into NR2A, NR2B, NR2C and NR2D. The NR3 subunit is divided into NR3A and NR3B.<sup>4,5</sup> Although NMDARs are also permeable to Na<sup>+</sup>, Ca<sup>+</sup>, and K<sup>+</sup> ions, and they most importantly are permeable to Ca<sup>++</sup> ions. The NR2 subunit of NMDARs provides an important level of receptor regulation. The NR2B subunit is primarily localized in the

dorsal horn of the spinal cord and in structures of the forebrain including the hippocampus, cortex and striatum, but not the cerebellum.<sup>6</sup> This restricted distribution of the NR2B subunit of NMDARs makes it a potential therapy target for the treatment of stroke, epilepsy, neuropathic pain, Parkinson's disease and Alzheimer's disease. Non-subtype selective NMDA receptor antagonists have undesirable side effects like dizziness, headaches, and hallucinations. These undesirable side effects can be overcome by the discovery of selective NR2B antagonists such as ifenprodil (**1**), Ro 25-6981 (**2**), traxoprodil (CP-101,606) (**3**) and besonprodil (CI-1041) (**4**).

Therefore, NR2B selective NMDAR antagonists are considered potential therapeutic agents that are advantageous over non subtype selective NMDA receptor antagonists such as memantine (**5**) and MK-801 (**6**).

Many new NR2B selective NMDAR antagonists have been discovered. The research efforts in the discovery of novel structural classes of NR2B subtype selective antagonists that do not fit the classical "ifenprodil-like" pharmacophore, have been described by Layton *et al.*<sup>7</sup> Classical "ifenprodil-like" pharmacophores containing NR2B selective NMDA antagonists have also been reported.<sup>8</sup> Other reviews on novel patented *N*-methyl-

Pharmaceutical Chemistry Department, Sinhgad Technical Education Society's Smt. Kashibai Navale College of Pharmacy, Savitribai Phule Pune University, Kondhwa (Bk.), Pune, India. E-mail: vishalzambre@gmail.com



D-aspartate receptor antagonists were also published.<sup>9,10</sup> A review on the structure–activity relationship and therapeutic applications of NR2B selective antagonists was published by Beinat *et al.*<sup>11</sup> Recently, novel NR2B selective NMDA receptor antagonists containing various heterocycles have been reported, such as aminocyclopentanes,<sup>12</sup> benzimidazole,<sup>4,13,14</sup> pyrazines,<sup>15</sup> tetrahydro-3-benzazepine-1,7-diols,<sup>16,17</sup> pyridine,<sup>18</sup> indole<sup>2,19–21</sup> benzoyl urea,<sup>22</sup> pyridinol,<sup>23</sup> quinolinone,<sup>24</sup> aminoquinoline,<sup>25</sup> benzamidine,<sup>26,27</sup> and carbamates.<sup>28</sup> Although many developments are taking place in the discovery of NR2B selective NMDA receptor antagonists, it is very difficult to discover clinically useful candidates. Study of previously reported compounds along with the application of some molecular modeling techniques, such as 2D-QSAR and 3D-QSAR,<sup>29–35</sup> would definitely help to speed up the discovery process. To the best of our knowledge, so far no such detailed QSAR study has been described on pyrazine and related derivatives as NR2B selective NMDA receptor antagonists. In this paper, we propose the first of its kind, 2D- and 3D-QSAR studies for pyrazine derivatives for their selectivity towards the NR2B subunit of a NMDA receptor. The objective of the present study is to elucidate the structural features of pyrazine derivatives, which may provide researchers new molecular insights for further optimization of 2,6-disubstituted pyrazine derivatives as NR2B selective NMDA receptor antagonists. Some important observations were also made during the study concerning the ligands and their interactions with the NMDA receptor.

## 2 Computational methods

### 2.1 Molecular docking

To investigate the molecular interaction mechanism for pyrazine and related derivatives with a NMDA receptor (PDB code 3QEL),

docking studies were performed using AutoDock 4.<sup>36</sup> In the RCSB protein data bank, around seven crystal structures for the NMDA receptor are available, *viz.* 3JPW, 3JPY, 4PE5, 3QEL, 3QEM, 4TLL and 4TLM. Out of these, 3JPW and 3JPY contain no co-crystallized ligand in them, whereas 4PE5, 3QEL, 4TLL and 4TLM contain ifenprodil as a co-crystallized ligand. In addition, the PDB 3QEM is co-crystallized with an ifenprodil derivative (Ro-25-6981). Among all these crystal structures, PDB 3QEL is the most well resolved, with a resolution 2.6 Å greater than that of the other crystal structures. Therefore, a highly resolved structure of PDB 3QEL was selected for the docking study in the present case. Validation of the docking study was performed by re-docking the co-crystallized ligand ifenprodil (1) into its binding pocket. A hydrogen bonding interaction was found between the N–H of the most active compound (18) and the Gln110 residue of the NMDA receptor. A similar hydrogen bonding interaction with the Gln110 residue of the NMDA receptor was also observed with the original co-crystallized ligand (1) in the 3D structure of the NMDA receptor. Moreover, the rmsd between the best predicted confirmation and the original confirmation was found to be 0.24 Å. Docking experiments were performed using a Lamarckian genetic algorithm. The lowest energy conformation of the most active compound 18, generated using docking studies, was selected as the bioactive conformation.

### 2.2 Biological data

The experimental biological data was chosen from a series of thirty five 2,6-disubstituted pyrazine derivatives (Table 1) reported by Brown *et al.* as NR2B selective NMDA receptor antagonists.<sup>15</sup> The present series of compounds has large structural variations and differences in the biological activity, with  $K_i$  values ranging from 12 to 8900 nM, making the series ideal for building reliable QSAR models. The biological activity, *i.e.*, the dependent parameter used in this study, was calculated as the negative logarithm of the molar unit of the  $K_i$  (nM) value.

### 2.3 Molecular modeling, descriptor calculation and alignment

A Pentium 4 personal computer with Windows XP workstation was used for all computations. Molecular modeling studies

Table 1 2,6-Disubstituted pyrazine and related derivatives with their biological data used in building and validating models

Compound	R1	R2	Biological data					
			NR2B bind $K_i$ (nM)	Experimental NR2B bind $pK_i$	Predicted NR2B bind $pK_i$			
					MLR	Residual	PLS	Residual
1 <sup>a</sup>	2-OMe, 4-F		27	7.57	7.2100	0.36	7.232	0.338
2 <sup>a</sup>	2-OMe, 4-F		53	7.28	6.9770	0.3030	7.0013	0.2787
3	2-OMe, 4-F		54	7.27	7.1815	0.0885	7.1526	0.1174
4	2-OMe, 4-F	-(CH <sub>2</sub> ) <sub>2</sub> OMe	214	6.67	6.5099	0.16010	6.5888	0.0812
5	2-OMe, 4-F	-CH <sub>2</sub> CH(CH <sub>3</sub> ) <sub>2</sub>	217	6.66	6.9770	-0.3170	7.0013	-0.3413
6 <sup>a</sup>	2-OMe, 4-F	-CH <sub>2</sub> CH <sub>3</sub>	255	6.59	6.5682	0.0218	6.6988	-0.1088
7 <sup>a</sup>	2-OMe, 4-F		429	6.37	6.7726	-0.4026	6.8501	-0.4801
8	2-OMe, 4-F		610	6.22	6.8165	-0.5965	6.8157	-0.5957
9	2-OMe, 4-F		3565	5.45	6.5428	-1.092	6.6885	-1.2385
10	2,3-Di-F		186	6.73	6.2476	0.4824	6.2731	0.4570
11	2,3-Di-F		341	6.47	6.2476	0.2224	6.2731	0.1970
12	2,3-Di-F		2979	5.53	6.0432	-0.5132	6.1218	-0.5918
13	2-F, 3-OMe		602	6.22	4.9215	1.2985	5.0122	1.2078
14	2-F, 3-OMe		5700	5.24	4.9928	0.2472	5.0223	0.2177
15	3-CN		1554	5.81	5.9271	-0.1171	5.8473	-0.0373
16	4-CN		1571	5.80	5.8249	-0.0249	5.7717	0.0283

Table 1 (Contd.)

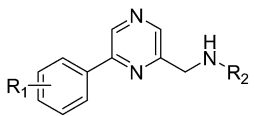
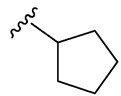
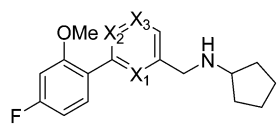
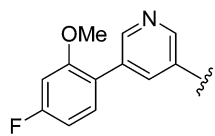
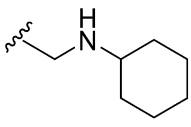
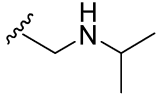
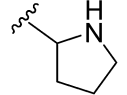
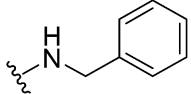
									
Biological data									
Compound	R1	R2	NR2B bind $K_i$ (nM)	Experimental NR2B bind $pK_i$	Predicted NR2B bind $pK_i$				
					MLR	Residual	PLS	Residual	
17	2-OMe, 4-OMe		1689	5.77	5.9710	-0.2010	5.8130	-0.0430	
									
Biological data									
Compound	X1	X2	X3	NR2B bind $K_i$ (nM)	Experimental NR2B bind $pK_i$	Predicted NR2B bind $pK_i$			
						MLR	Residual	PLS	Residual
18	C	C	N	12	7.92	7.1815	0.7385	7.1526	0.7674
19 <sup>a</sup>	C	C	C	48	7.32	7.1815	0.1385	7.1526	0.1674
20	N	C	C	66	7.18	7.1815	-0.0015	7.1526	0.0274
21	C	N	C	8900	5.05	7.1815	-2.1315	7.1526	-2.1026
									
Biological data									
Compound	Side chain	NR2B bind $pK_i$ (nM)	Experimental NR2B bind $pK_i$	Predicted NR2B bind $pK_i$					
				MLR	Residual	PLS	Residual		
22 <sup>a</sup>		112	6.95	7.2837	-0.3337	7.2282	-0.2782		
23 <sup>a</sup>		266	6.65	6.7726	-0.1226	6.8501	-0.2001		
24 <sup>a</sup>		440	6.36	6.4077	-0.0477	6.5132	-0.1532		
25		457	6.34	6.9188	-0.5788	6.8914	-0.5514		

Table 1 (Contd.)

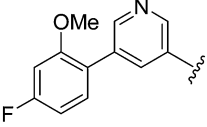
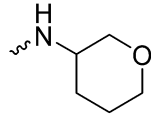
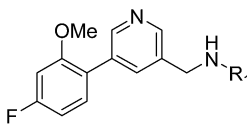
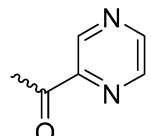
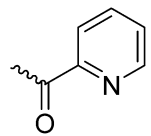
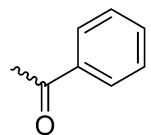
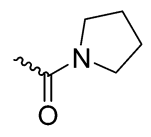
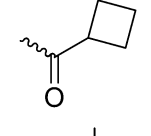
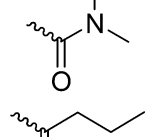
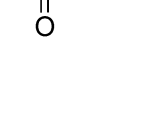
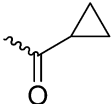
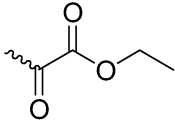
							
Biological data							
Compound	Side chain	NR2B bind $pK_i$ (nM)	Experimental NR2B bind $pK_i$	Predicted NR2B bind $pK_i$			
				MLR	Residual	PLS	Residual
26		705	6.15	6.0867	0.0633	6.1420	0.0080
							
Biological data							
Compound	R1	NR2B bind $K_i$ (nM)	Experimental NR2B bind $pK_i$	Predicted NR2B bind $pK_i$			
				MLR	Residual	PLS	Residual
27		80	7.1	6.7582	0.3418	6.7058	0.3942
28		116	6.94	6.7582	0.1818	6.7058	0.2342
29		450	6.35	6.7582	-0.4082	6.7058	-0.3558
30		268	6.57	6.6560	-0.0860	6.6301	-0.0601
31		340	6.47	6.5538	-0.0838	6.5545	-0.0845
32		420	6.38	6.4516	-0.0716	6.4789	-0.0989
33		468	6.33	6.3494	-0.0194	6.4032	-0.0732

Table 1 (Contd.)

Compound	R1	Biological data		Predicted NR2B bind $pK_i$			
		NR2B bind $K_i$ (nM)	Experimental NR2B bind $pK_i$	MLR	Residual	PLS	Residual
34		595	6.23	6.4516	-0.2216	6.4789	-0.2489
35		778	6.11	5.9262	0.1838	5.9564	0.1536

<sup>a</sup> Test set compounds.

were carried out using the molecular modeling software package VLife Molecular Design Suite (VLifeMDS, 4.3).<sup>37</sup> For 2D-QSAR studies, the structures of compounds were sketched using the ChemDraw Ultra 8.0 program and were converted into 3D structures using the Vlife 2D to 3D conversion tool. For 3D-QSAR, structures were constructed using coordinates of the docked conformation of compound **18**. The energy minimization of the structures was performed to remove close atom contacts using a Merck molecular force field (MMFF) with Gasteiger–Marsili charges. Then, 1000 cycles were run until a convergence criterion of 0.001 was achieved. These geometrically optimized structures were used for building 2D- and 3D-QSAR models. Two-dimensional descriptors (physicochemical and topological, 245 in number) and three-dimensional field descriptors (steric, electronic and hydrophobic) were calculated using the VLifeMDS software for 2D- and 3D-QSAR studies, respectively.

For 3D-QSAR, the field descriptors were calculated with cutoffs of 10.0 kcal mol<sup>-1</sup> for electrostatic and 30.0 kcal mol<sup>-1</sup> for steric using Gasteiger–Marsili charges (Gasteiger *et al.*, 1980).<sup>38</sup> The dielectric constant was set to 1.0, considering the distance-dependent dielectric function. A carbon atom with charge 1.0 was selected as the probe atom. All compounds were superimposed on the most active compound **18** by template-based, atom-based and docked-based alignment methods. For template-based and atom-based alignments, the energy minimized structure of most active compound **18** obtained by the above-mentioned method was used. For docked-based alignment, the docked pose of compound **18** was used and offered the lowest energy bioactive conformation. It was thus selected as a template for further alignment of all compounds using the atom-based alignment method. Results of the best alignment (*i.e.* docked-based alignment) are included in this report. The

atoms that were considered for the alignment are marked with an asterisk (\*) in Fig. 1. A superimposition of all of the compounds on the template is shown in Fig. 2.

#### 2.4 Training and test set selection

The thirty five compounds were divided into training (26) and test (09) compounds. The test set contains 25–35% of all the compounds to establish a reliable QSAR model.<sup>39</sup> In the training and test sets, the most potent, moderately active and poorly active compounds were included to spread the biological activity range. The similarity of the distribution pattern of the molecules in the generated training and test sets was judged

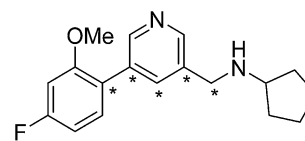


Fig. 1 Template and atoms considered for the alignment.

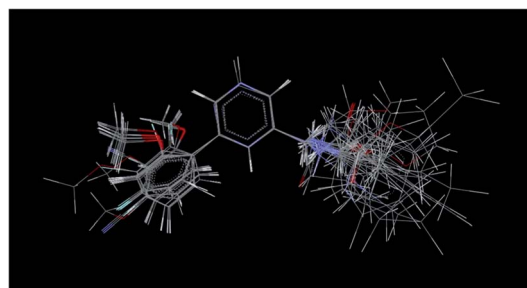


Fig. 2 Alignment of all compounds.

based on statistical parameters like mean, maximum, minimum and standard deviation.

## 2.5 2D-QSAR

**Multiple linear regression (MLR) method.** MLR is a standard method for finding a linear relationship between dependent and independent variables. It is also called ordinary least square regression analysis (OLS). The values of the regression coefficient were estimated by applying least squares curve fitting method. For reliable results, the parameters were set such that the number of descriptors were not more than  $1/5^{\text{th}}$  the number of data points or molecules in the training set.

The general regression equation for the MLR method is as follows:

$$Y = b_1 \times x_1 + b_2 \times x_2 + b_3 \times x_3 + c \dots \quad (1)$$

where  $Y$  is the dependent variable, ' $b$ 's are regression coefficients for the corresponding independent variable ' $x$ ', and ' $c$ ' is a regression constant or intercept.<sup>40,41</sup>

**Partial least square regression (PLSR) method.** Herman Wold developed the PLS method for finding the relationship between the properties of a molecule and its structure. This method is used to relate the biological activity with the number of descriptors (molecular structural descriptors). This method relates the matrix  $Y$  of the dependent variables (Biological activity) to matrix  $X$  of the independent variable, *i.e.*, molecular structure descriptors.<sup>42,43</sup> The objectives of the PLS method are to estimate the  $X$  and  $Y$  data matrices and to maximize the correlation between these matrices. Number of PLS components are extracted and a regression equation for correlation of  $Y$  variable with  $X$  matrix is created. PLS determines latent variables which correlate maximum with the biological activity. ( $Y$  dependent variable).<sup>44</sup>

## 2.6 3D-QSAR

**$k$ -Nearest neighbor molecular field analysis ( $k$ NN-MFA).** The  $k$ NN method is a simple distance learning approach wherein a classification of data can be achieved according to the majority of its  $k$ NN in the training set. In the  $k$ NN-MFA method, several models were generated for selected members of the training and test sets. After training and test set data selection, the  $k$ NN methodology is applied to calculated descriptors over the grid.<sup>45</sup> The steric and electrostatic interaction energies are computed at lattice points of the grid using a methyl probe of charge +1. These interaction energy values are considered for relationship generation and utilized as descriptors to decide the proximity between molecules.

**SW-FB variable selection method.** Models were generated using the  $k$ NN-MFA SW-FB variable selection method with cross-correlation limit set to 1.0 and term selection criteria as  $q^2$ .  $F$ -test 'in' and 'out' was set to 4.0 and 3.99, respectively. Other parameters such as variance cut-off were set to 0 kcal mol<sup>-1</sup> Å<sup>-1</sup> and scaling to an auto scaling mode.

**$k$ NN-MFA with simulated annealing.** Another model was also developed using simulated annealing as a variable selection

method. In simulated annealing, the system is heated to a certain high temperature and cooled to a lower temperature. The cross-correlation limit was set to 1, maximum temperature as 100, minimum temperature as 0.01, perturbation limit as 1, and decrease in temperature by 2. Iteration at the given temperature was kept at 5 and seed at 0. The term selection criteria were  $q^2$ , cut-off variance of 0, scaling set to autoscaling, number of maximum neighbors as 5 and minimum neighbors as 2.

## 2.7 Model validation

**Internal validation.** Developed 2D- and 3D-QSAR models were validated to establish their internal stability and predictive ability. Cross-validation is one of the most extensively used methods for internal validation, which involves a leave-one-out ( $q^2$  LOO) cross-validation. To calculate  $q^2$ , every compound was removed from the dataset once; dataset was further divided into subset of equal size. The activity of an excluded molecule was predicted using the developed model by the dataset of the remaining molecules. This procedure is repeated until the molecules have each been eliminated once. An equation for the calculation of  $q^2$  is as follows:

$$q^2 = 1 - \frac{\sum (Y_{\text{pred}} - Y_{\text{act}})^2}{\sum (Y_{\text{act}} - Y_{\text{mean}})^2} \quad (2)$$

where,  $Y_{\text{pred}}$  and  $Y_{\text{act}}$  are the predicted and actual activity of molecules in the training set, respectively.  $Y_{\text{mean}}$  is the average activity of all molecules in the training set.

**External validation.** The predictive power of the QSAR models was validated by using an external test set (inhibitors marked with 'a' in Table 1). The inhibitors in the test set were given exactly the same pre-treatment as the inhibitors in the corresponding training set. The predictive ability of the model was calculated using the following equation:

$$\text{pred}_r^2 = 1 - \frac{\sum (Y_{\text{pred(} \text{Test)}} - Y_{\text{Test}})^2}{\sum (Y_{\text{Test}} - Y_{\text{training}})^2} \quad (3)$$

where,  $Y_{\text{pred(} \text{Test)}}$  and  $Y_{\text{Test}}$  are the predicted and observed activities of test set compounds, respectively, and  $Y_{\text{training}}$  is the mean activity of all molecules in the training set. The graph of the observed activity *versus* the predicted activity ( $-\log K_i$ ) indicates that the model is statistically significant. Golbraikh and Tropsha reported that the actual predictive ability of a QSAR model can only be estimated using an external test set of compounds that were not used for building the model.<sup>46</sup>

**Randomization test.** The robustness of the QSAR models was evaluated by a  $Y$ -randomization test. The robustness of the models for the training sets was examined by comparing these models to those derived for random datasets. Random sets were generated by rearranging the biological activities of the molecules in the training set. The significance of the models was derived based on the calculated  $Z$  score.

The  $Z$  score value was calculated using the following formula:

$$Z \text{ score} = \frac{(q_{\text{org}}^2 - q_{\text{a}}^2)}{q_{\text{std}}^2} \quad (4)$$

where  $q_{\text{org}}^2$  is the  $q^2$  value calculated for the actual data set,  $q_{\text{a}}^2$  is the average  $q^2$ , and  $q_{\text{std}}^2$  is the standard deviation of  $q^2$ , calculated for various iterations using different random data sets.<sup>47</sup>

**Other metrics to judge predictive power of models.** The predictive power of all four models was judged by a wide range of validation parameters suggested by Golbraikh and Tropsha.<sup>46,48</sup> Additional validation parameters, such as  $R_{\text{m}}^2$ -(training),  $R_{\text{m}}^2$ -(test) and  $R_{\text{m}}^2$ -(overall) for internal, external and overall validation, respectively, were computed to check the quality of the models.<sup>49,50</sup>

Other validation parameters, such as  $\{[(r^2 - r_0^2)/r^2] < 0.1 \text{ or } [(r'^2 - r_0'^2)/r'^2] < 0.1\}$ , were also determined, where  $r_0^2$  and  $r_0'^2$  are correlation coefficients for the regression line without intercept for actual *versus* predicted and predicted *versus* actual activity, respectively. The  $k$  and  $k'$ , accuracy, sensitivity, specificity, positive prediction value (PPV), negative prediction value (NPV) and Matthew's correlation coefficient (MCC) values were calculated for the models. A new validation parameter, % of residual value (% residuals), reported by Zambre *et al.* was also evaluated for the QSAR models.<sup>31</sup>

## 3 Results and discussion

### 3.1 2D-QSAR

2D-QSAR analysis initially was performed on reported compounds by dividing them into a training set (26) and test set (9). The model obtained thus was observed to be statistically least significant. Based on the statistical analysis, compounds **9**, **13** and **21** showed residuals more than 1 unit in value. In the series of compounds, compound **9** was the only compound with 2-pyridyl aromatic substitution at the right hand side chain in the data set. This structural uniqueness of compound **9** may lead to the higher residual value. Compound **14**, with similar functional groups at the R1 and R2 positions to that of compound **13**, was comparatively well predicted, and for compounds **18–21**, heteroatoms at the X1, X2 and X3 positions in the central aromatic ring are bioisosteres of each other. Among these four compounds, compound **21** was over-predicted. Hence, it is most likely that outliers (compounds **13** and **21**) might be a result of an experimental error that could happen when analyzing a large data set.<sup>51</sup> After removing these compounds, a newly formed training set (24) and test set (8) were assessed for similarity of the distribution patterns (Table 2) of the molecules. Results showed that the max of the test set is less than the max of the training set and the min of the test set is greater than that of the training, which is a prerequisite for further QSAR study. The minimum 'NR2B binding affinity' (biological activity) of the test set is greater than the minimum activity of the training set, and the maximum activity of the test set is less than the maximum activity of the training set. This indicates that the test set is within the activity domain of the training set. A higher mean value of the test set than of the training set indicates the presence of relatively more potent compounds, as compared to inactive ones.

**Table 2** Statistical parameters for biological activity distribution in training and test set for 2D-QSAR

Parameters	Training set	Test set
Maximum	7.9200	7.5700
Minimum	5.2400	6.3400
Mean	6.4275	6.8838
Standard deviation	0.5953	0.4665

MLR and PLS analysis with these newly formed data sets showed more statistically significant results (Table 3).

#### Model-1 (MLR).

$$-\log(K_i) = 0.5250(\text{T}_2\text{F}_7) - 0.3649(\text{T}_\text{C}_\text{O}_7) + 0.1022(\text{T}_\text{T}_\text{T}_6) + 3.6784$$

where  $n = 24$ , degree of freedom = 20,  $r^2 = 0.7397$ ,  $q^2 = 0.6258$ ,  $F$  test = 18.9418,  $r^2$  se = 0.3257,  $q^2$  se = 0.3906,  $\text{pred}_r^2 = 0.6592$ ,  $\text{pred}_r^2$  se = 0.3940

#### Model-2 (PLS).

$$-\log(K_i) = 0.5770\text{T}_2\text{F}_7 - 0.3369\text{T}_\text{C}_\text{O}_7 + 0.0756\text{T}_\text{T}_\text{T}_6 + 4.1833$$

where  $n = 24$ , degree of freedom = 21,  $r^2 = 0.7298$ ,  $q^2 = 0.6273$ ,  $F$  test = 28.3641,  $r^2$  se = 0.3239,  $q^2$  se = 0.3804,  $\text{pred}_r^2 = 0.6573$ ,  $\text{pred}_r^2$  se = 0.3951.

The above QSAR models have  $r^2$  correlation coefficients of 0.7397 and 0.7298 for MLR and PLS, respectively. The predictive ability of generated QSAR models was evaluated by  $q^2$  (0.6258 for MLR and 0.6273 for PLS) employing the LOO method. The  $F$  value indicates the ratio of variance explained by the models and variance due to error in regression. The higher  $F$  value of 28.3641 for the PLS model indicates that the model is statistically significant. Values of  $R_{\text{m}}^2$ (training),  $R_{\text{m}}^2$ (test),  $R_{\text{m}}^2$ (overall),  $\{[(r^2 - r_0^2)/r^2] < 0.1 \text{ or } [(r'^2 - r_0'^2)/r'^2] < 0.1\}$ ,  $k$  and  $k'$ , accuracy, sensitivity, specificity, PPV, NPV, MCC and % residuals were calculated and found to be within acceptable limits. These validation parameter values are summarized in Table 3. A graphical representation of the accuracy, sensitivity, PPV, NPV and MCC values is given in Fig. 3. Graphs of predicted *versus* experimental activities with intercept and without intercept for training set and test set compounds are shown in Fig. 4. The 'fitness plot', *i.e.*, plot of predicted *versus* experimental activity of training and test set compounds (Fig. 4) for each model, shows that the built models are statistically significant. This suggests that the models were correctly trained and accurately predict the activity of external test set compounds. Residual values for each 2D-QSAR model are presented in Table 1. The percentage contributions of descriptors in deriving both of the significant QSAR models are shown in Fig. 5.

**Interpretation of 2D-QSAR models.** MLR and PLS models indicate positive contributions of Baumann's alignment independent (AI) topological descriptors such as  $\text{T}_2\text{F}_7$  &  $\text{T}_\text{T}_\text{T}_6$  and a negative contribution of  $\text{T}_\text{C}_\text{O}_7$ , as shown in Fig. 5. The AI descriptors (Table 4) consider the topology of the molecule,



Table 3 Summary of validation parameters for 2D-QSAR models

Statistical parameter	Model-1 (MLR)	Model-2 (PLS)
$N^a$	24	24
DF <sup>b</sup>	20	21
$r^{2c}$	0.7397	0.7298
$q^{2d}$	0.6258	0.6273
F-test <sup>e</sup>	18.9418	28.3641
$r^2$ se <sup>f</sup>	0.3257	0.3239
$q^2$ se <sup>g</sup>	0.3906	0.3804
Pred_ $r^{2h}$	0.6592	0.6573
Pred_ $r^2$ se <sup>i</sup>	0.3940	0.3951
Z score_ CV <sup>j</sup>	5.80396	5.80396
Best_rand- $q^{2k}$	0.18513	0.08517
$\alpha$ _rand_ CV <sup>l</sup>	0.00000	0.00000
$R^{2m}$	0.6619	0.6849
$R_0^{2n}$	0.2545	-0.246
$k^o$	0.9998	1.0058
$R_0'^{2p}$	0.6462	0.6119
$k'^q$	0.9988	0.9927
$(R^2 - R_0'^2)/R^2$	0.0231	0.1
$R_m^2$ (test)	0.579	0.4999
$R_m^2$ (training)	0.6766	0.6758
$R_m^2$ (overall)	0.7205	0.7125
Absolute sum of observed $pK_i$ (test set)	55.1733	55.5263
Absolute sum of residuals (test set)	1.7299	2.0045
% Residual (test set)	3.14%	3.61%
Absolute sum of observed $pK_i$ (training set)	154.7708	154.6387
Absolute sum of residuals (training set)	5.9504	5.7651
% Residual (training set)	3.844%	3.73%

<sup>a</sup> Number of molecules. <sup>b</sup> Degree of freedom. <sup>c</sup> Non-cross validated  $r^2$ . <sup>d</sup> Cross validated  $q^2$ . <sup>e</sup> Fisher test value. <sup>f</sup> Std error for non-cross validated  $r^2$ . <sup>g</sup> Std error for non-cross validated  $q^2$ . <sup>h</sup> Predicted  $r^2$ . <sup>i</sup> Std error for predicted  $r^2$ . <sup>j</sup> probability of significance of randomization test. <sup>k</sup> highest  $q^2$  value in the randomization test. <sup>l</sup> Statistical significance of randomization test. <sup>m</sup> Correlation coefficient derived from predicted  $pK_i$  of test set compounds. <sup>n</sup> Correlation coefficient for the regression line passing through origin for experimental vs. predicted activities. <sup>o</sup> Slope for regression line passing through origin obtained from actual vs. predicted activities. <sup>p</sup> Correlation coefficient for the regression line passing through origin for predicted vs. experimental activities. <sup>q</sup> Slope for regression line passing through origin obtained from predicted vs. experimental activities.

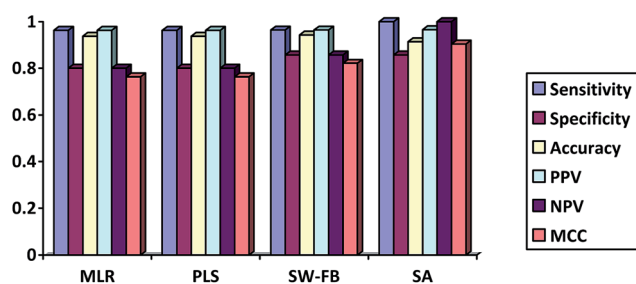


Fig. 3 Validation of models using some extra parameters.

atom type and bond. First, 'T-attribute' thoroughly characterizes the topology of the molecule. Second, the attribute is atom type, denoted with atom symbol. Third, the attribute is assigned to a double or triple bond for the atoms. The descriptor T\_2\_F\_7 was found to be directly proportional to the biological activity variation, which suggests that an electronegative fluorine atom is detrimental for the activity. The most active compound **18** ( $K_i$  12 nM) possesses an electronegative fluorine group at the 4<sup>th</sup> position of the phenyl ring, whereas the less active compound

**17** ( $K_i$  1689 nM) possesses an electropositive methoxy group at the 4<sup>th</sup> position of the phenyl ring. This indicates an importance of electronegative substitutions at the fourth position of the phenyl ring. The fluorine group is lipophilic in character and cyano group is hydrophilic in character. Hence, compound **16** ( $K_i$  1571 nM), possessing a hydrophilic cyano group at the 4<sup>th</sup> position of the phenyl ring leads to decrease in activity. These observations revealed that the 4<sup>th</sup> position of the phenyl ring should hold more electronegativity and lipophilic character for better NR2B subunit selectivity of NMDA receptors. The T\_T\_T\_6 descriptor indicates the count of the number of any atom (single, double or triple bonded) separated from any other atom (single or double bonded) by 6 bond distances in a molecule. This clearly states that a six-membered aromatic ring is an important requirement for the activity. The descriptor T\_C\_O\_7 is the count of any pair of a carbon atom and an oxygen atom separated by 7 bonds. Here, the -OCH<sub>3</sub> group present at the 2<sup>nd</sup> position of the phenyl ring, was found to be effective for retaining NR2B selectivity of NMDA receptors. It was stated that the 2-MeO and 4-F aromatic substitution pattern is preferred for optimum NR2B binding affinity.<sup>15</sup> The results of the present study corroborates the above observation, as the 2D-

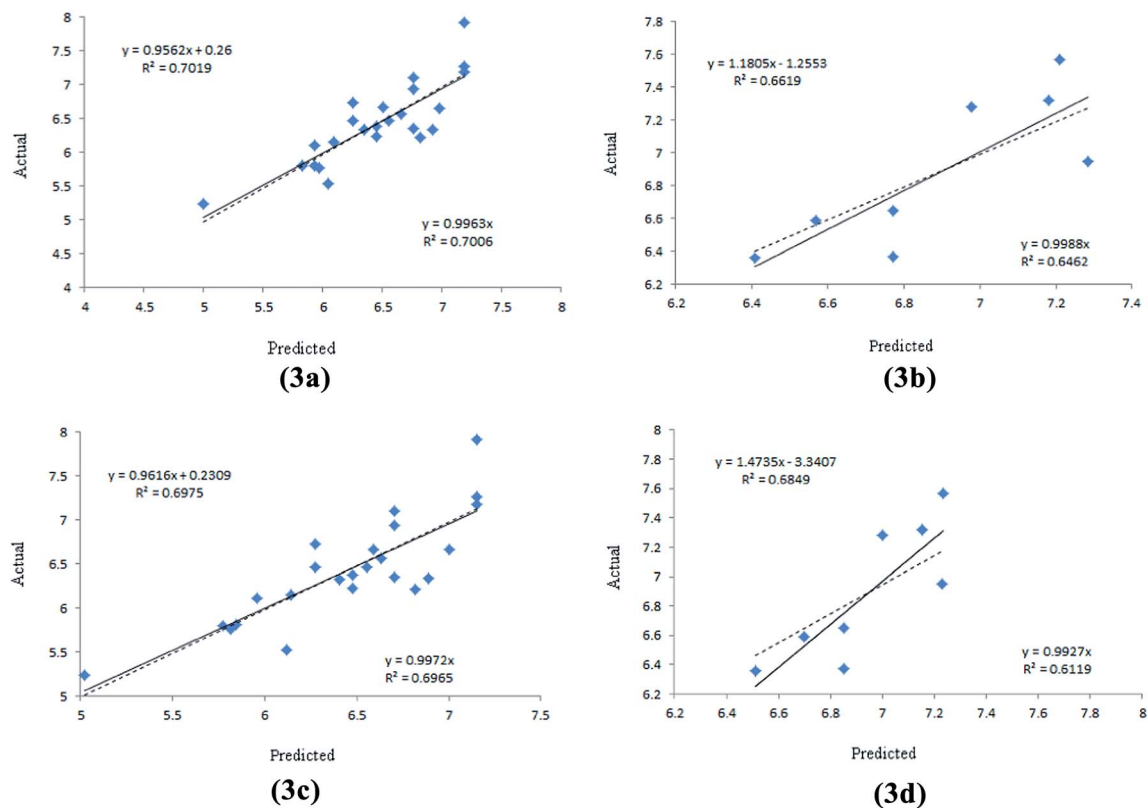


Fig. 4 Graph of predicted vs. experimental activities from the best predictive 2D-QSAR models; MLR training set (3a), test set (3b) and PLS training set (3c), test set (3d) models.

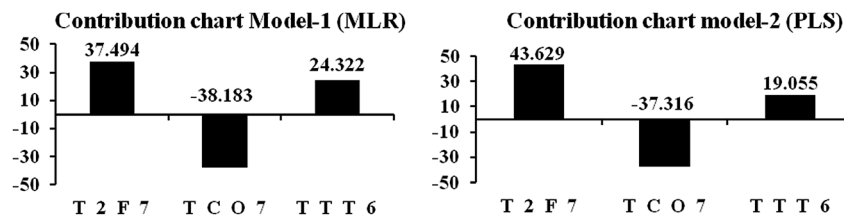


Fig. 5 Contribution chart for MLR and PLS models.

Table 4 Significance of 2D molecular descriptors used in QSAR study

Descriptors	Significance
T <sub>2</sub> F <sub>7</sub>	The count of the number of double bonded atoms ( <i>i.e.</i> any double bonded atom, T <sub>2</sub> ) separated from a fluorine atom by 7 bonds in a molecule
T <sub>C</sub> O <sub>7</sub>	The count of the number of carbon atoms (single, double or triple bonded) separated from any oxygen atom (single or double bonded) by 7 bond distances in a molecule
T <sub>T</sub> T <sub>6</sub>	The count of the number of any atom (single, double or triple bonded) separated from any atom (single or double bonded) by 6 bond distances in a molecule

QSAR alignment independent descriptors, such as T<sub>C</sub>O<sub>7</sub> (oxygen atom of methoxy group at 2<sup>nd</sup> position of aromatic ring) and T<sub>2</sub>F<sub>7</sub> (fluorine atom at 4<sup>th</sup> position of aromatic ring), were found to be well correlated with NR2B binding  $K_i$

values. Compound 14 ( $K_i$ : 5700 nM) does not meet the above substitution pattern, thus it was found to be the least active compound in the series.

### 3.2 Molecular docking

The most active compound **18** was docked into the active site of the NMDA receptor (PDB Id: 3QEL) using AutoDock 4. The binding energy for this complex was found to be  $-9.0$ . The simple energy minimised structure may not be the bioactive conformation. Therefore, the generated lowest energy conformation using this validated docking study was used as a bioactive conformation for 3D-QSAR studies. Analysis of docking poses revealed that compound **18** fitted well into the active site of the NMDA receptor (Fig. 6). The  $\pi$ -cation,  $\pi$ -sigma and H-bond interactions with some conserved residues were observed. Strong H-bond interactions were observed between the N-H of compound **18** and Gln110 residue, as well as between the oxygen atom of the methoxy group and the Arg115 residue of the NMDA receptor. Moreover, compound **18** forms interactions with some hydrophobic amino acid residues like Tyr109 and Leu135 present on the GluN1b portion and Ile111, Phe176, and Pro177 present on the GluN2B portion of the NMDA receptor. The docking pose and orientation of the most active compound **18** revealed that the left side pocket in the NMDA receptor still has room to accommodate more bulky groups, thus it may be possible that fused aromatic or hetero-aromatic rings may be capable of accommodating this site (Fig. 6) for improved selectivity towards the NR2B subunit of the NMDA receptor. Based on the key residual interactions, we propose a putative structure-based pharmacophore model containing the features of two aromatic rings, one H-bond acceptor and one H-bond donor, as shown in Fig. 7. The distances among the pharmacophoric features using the bioactive confirmation of the most active compound **18** obtained from the docking study were calculated as shown in Fig. 7. This pharmacophore model could be useful for virtual screening of commercially available chemical databases to identify new hits, and 2D- and 3D-QSAR models could be used to predict the activities of identified hits.

### 3.3 3D-QSAR models

3D-QSAR studies were performed on 35 previously reported NR2B selective NMDA receptor antagonists and divided into training set (26) and test set (9). The training set and test set were assessed for similarity of the distribution patterns (Table 5) of the molecules.

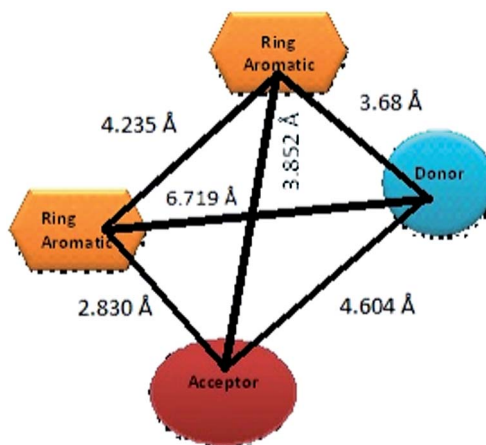


Fig. 7 Proposed putative pharmacophore model from docking study.

3D-QSAR models were built using *k*NN-MFA SW-FB and SA methods. Steric (S), electrostatic (E) and hydrophobic (H) fields were calculated for each compound in the series. The statistical quality of the models was judged based on the value of  $q^2$ , the internal predictive ability of model, and on the external predictive ability ( $\text{pred}_r^2$ ) of the model to predict the activity of the test set compounds. The above two methods offered statistically significant models (Model-3 and 4) and correlate steric and electronic descriptors with the NR2B selectivity of pyrazine derivatives. 3D data points generated around the pyrazine nucleus were used to study the steric and electrostatic requirements for NR2B selective NMDA antagonists.

Table 5 Statistical parameters for biological activity distribution in training and test set for 3D-QSAR

Parameters	Training set	Test set
Maximum	7.9200	7.5700
Minimum	5.0500	5.5300
Mean	6.4446	6.4989
Standard deviation	0.6523	0.6658

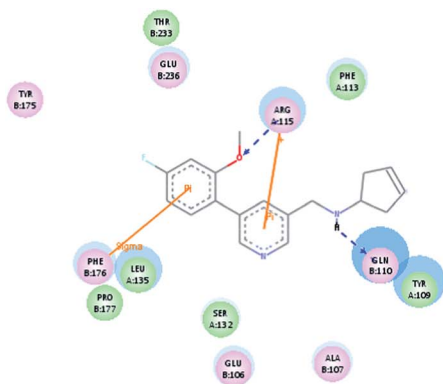


Fig. 6 Docking interaction model and orientation of compound **18**.

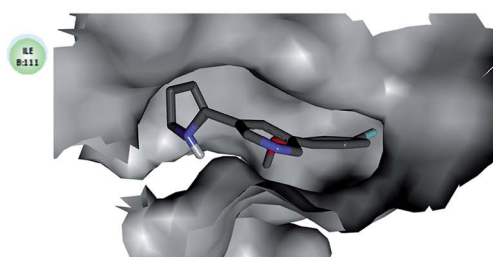


Table 6 Summary of validation parameters for 3D-QSAR models

Validation parameters	kNN-MFA SW-FB (Model-3)	kNN-MFA SA (Model-4)
kNN <sup>a</sup>	2	2
N <sup>b</sup>	26	26
DF <sup>c</sup>	22	20
q <sup>2d</sup>	0.8103	0.7799
q <sup>2</sup> se <sup>e</sup>	0.2904	0.3128
Pred_r <sup>2f</sup>	0.6556	0.7097
pred_r <sup>2</sup> se <sup>g</sup>	0.3655	0.3355
R <sup>2m</sup>	0.8206	0.7899
R <sub>0</sub> <sup>2n</sup>	0.8131	0.7844
k <sup>o</sup>	0.9125	0.9297
(R <sup>2</sup> - R <sub>0</sub> <sup>2</sup> )/R <sup>2</sup>	0.0091	0.0069
R <sub>0</sub> <sup>2p</sup>	0.8105	0.7678
k <sup>q</sup>	0.9985	0.9911
(R <sup>2</sup> - R <sub>0</sub> <sup>2</sup> )/R <sup>2</sup>	0.0123	0.0279
R <sub>m</sub> <sup>2</sup> (test)	0.5216	0.6406
R <sub>m</sub> <sup>2</sup> (training)	0.7495	0.7313
R <sub>m</sub> <sup>2</sup> (overall)	0.7461	0.7361
Absolute sum of observed pK <sub>i</sub> (test set)	57.55	57.55
Absolute sum of residuals (test set)	2.61	1.97
% Residual (test set)	4.53	3.42
Absolute sum of observed pK <sub>i</sub> (training set)	168.5	168.5
Absolute sum of residuals (training set)	6.43	6.68
% Residual (training set)	3.81	3.96
Descriptors		E_23
	E_202	S_1033
	E_966	E_1044
	E_1030	S_1071
		S_1118

<sup>a</sup> k nearest neighbour. <sup>b</sup> Number of molecules. <sup>c</sup> Degree of freedom. <sup>d</sup> Cross validated q<sup>2</sup>. <sup>e</sup> Std error for cross validated q<sup>2</sup>. <sup>f</sup> Predicted r<sup>2</sup>. <sup>g</sup> Std error for predicted r<sup>2</sup>.

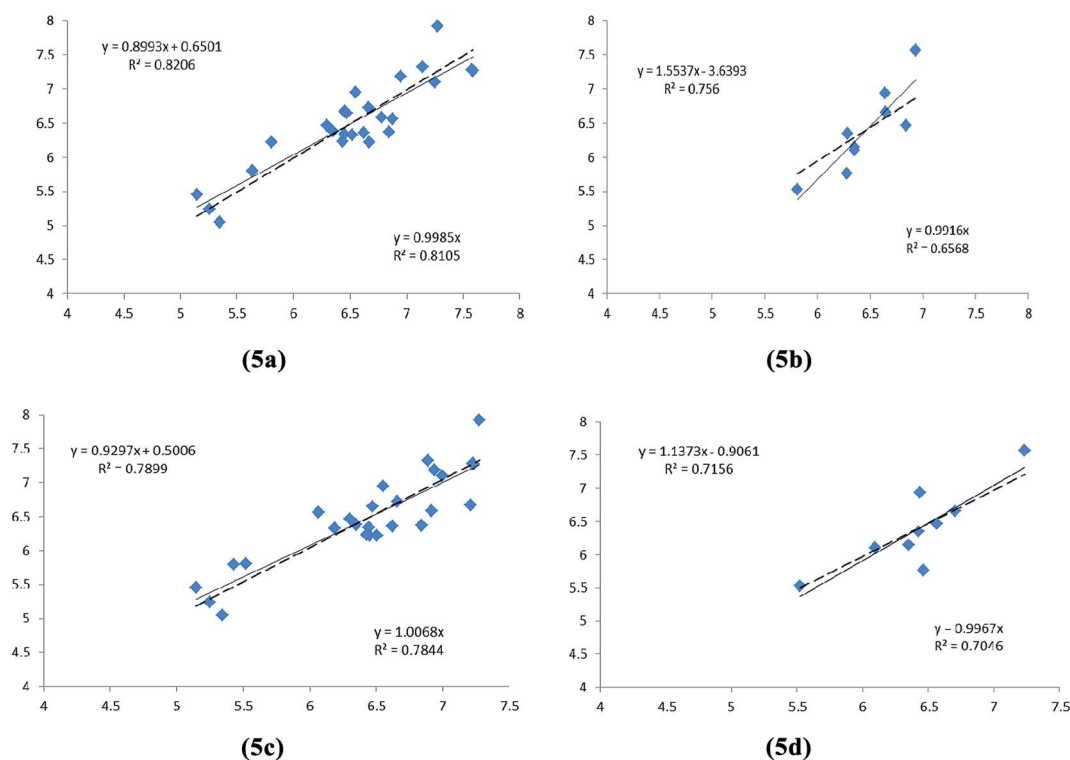


Fig. 8 Graph of predicted vs. experimental activities from the best predictive 3D-QSAR models; SW-FB training set (5a), test set (5b) and SA training set (5c), test set (5d) models.

**Model-3 (SW-FB).**

$$\log(K_i) = E_{202} (0.0380 \ 0.0390) E_{966} -0.1130 -0.0820 E_{1030} \\ 0.4580 \ 0.4750$$

**Model-4 (SA).**

$$-\log(K_i) = E_{23} (0.0040 \ 0.0040) S_{1033} (-0.0350 -0.0290) \\ E_{1044} (-0.0290 -0.0190) S_{1071} (-0.0050 \\ -0.0040) S_{1118} (-0.0750 -0.0550)$$

**Model validation.** Model-3 and 4 were found to be statistically significant, with the value of the LOO cross-validation squared correlation coefficient  $q^2 = 0.8103$  and  $0.7799$ , respectively, which suggested goodness of the predictions. Both models have high predictive ability with predictive squared correlation coefficients ( $\text{pred}_r^2$ ) of  $0.6556$  (Model-3) and  $0.7097$

(Model-4), which are in agreement with the accepted criteria of more than  $0.4$ . Model-3 and 4 were found to correlate with steric and electronic interaction fields, indicating that these two descriptors are the key structural features for the optimization of pyrazine and related derivatives for use as NR2B subunit selective NMDA receptor antagonists.

Values of  $R_m^2(\text{training})$ ,  $R_m^2(\text{test})$ ,  $R_m^2(\text{overall})$ ,  $\{[(r^2 - r_0^2)/r^2] < 0.1$  or  $[(r'^2 - r_0'^2)/r'^2] < 0.1\}$ ,  $k$  and  $k'$ , accuracy, sensitivity, specificity, PPV, NPV, MCC and % residuals were calculated and found to be within acceptable limits. All the validation parameter values are summarized in Table 6.

The predicted *versus* experimental correlation graph for the training and test sets for Model-3 and 4 are depicted in Fig. 8. Residual values for the SW-FB and SA 3D-QSAR models are given in Table 7. Between these two models, Model-3 (SW-FB) displayed higher prediction ability both in regular cross-validation and in the prediction of the test compounds, whereas Model-4 (SA) offers sufficient 3D points to study the

Table 7 Residuals of 3D-QSAR

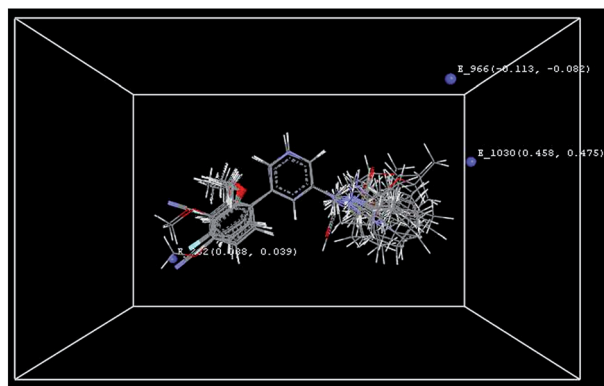
Compound	NR2B binding $K_i$ (nM)	$-\log(K_i)$ (observed)	Predicted NR2B binding $pK_i$			
			$k$ NN MFA SW-FB	Residuals	$k'$ NN MFA SA	Residuals
1 <sup>a</sup>	27	7.57	6.931818	0.638182	7.231062	0.338938
2	53	7.28	7.575945	-0.295945	7.224959	0.055041
3	54	7.27	7.584706	-0.314706	7.230381	0.039619
4	214	6.67	6.452823	0.217177	7.209869	-0.53987
5 <sup>a</sup>	217	6.66	6.648324	0.011676	6.70122	-0.04122
6	255	6.59	6.779736	-0.189736	6.916679	-0.32668
7	429	6.37	6.842204	-0.472204	6.840003	-0.47000
8	610	6.22	5.804974	0.415026	6.450573	-0.23057
9	3565	5.45	5.145978	0.304022	5.144824	0.305176
10	186	6.73	6.664754	0.065246	6.660224	0.069776
11 <sup>a</sup>	341	6.47	6.837207	-0.367207	6.563657	-0.09366
12 <sup>a</sup>	2979	5.53	5.805441	-0.275441	5.521352	0.008648
13	602	6.22	6.668226	-0.448226	6.504833	-0.28483
14	5700	5.24	5.254856	-0.014856	5.249304	-0.00930
15	1554	5.81	5.635744	0.174256	5.521223	0.288777
16	1571	5.80	5.634587	0.165413	5.430364	0.369636
17 <sup>a</sup>	1689	5.77	6.278748	-0.508748	6.459184	-0.68918
18	12	7.92	7.274946	0.645054	7.274963	0.645037
19	48	7.32	7.138975	0.181025	6.887164	0.432836
20	66	7.18	6.949905	0.230095	6.93614	0.24386
21	8900	5.05	5.346469	-0.296469	5.34483	-0.29483
22	112	6.95	6.549344	0.400656	6.550134	0.399866
23	266	6.65	6.471161	0.178839	6.473628	0.176372
24	440	6.36	6.620902	-0.260902	6.620445	-0.26045
25	457	6.34	6.453336	-0.113336	6.444119	-0.10412
26 <sup>a</sup>	705	6.15	6.349979	-0.199979	6.348554	-0.19855
27	80	7.1	7.2521	-0.1521	6.997884	0.102116
28 <sup>a</sup>	116	6.94	6.64113	0.29887	6.4348	0.50520
29 <sup>a</sup>	450	6.35	6.280839	0.069161	6.422796	-0.0728
30	268	6.57	6.875065	-0.305065	6.067154	0.502846
31	340	6.47	6.293945	0.176055	6.302404	0.167596
32	420	6.38	6.35067	0.02933	6.349836	0.030164
33	468	6.33	6.520304	-0.190304	6.188163	0.141837
34	595	6.23	6.431879	-0.201879	6.426497	-0.19650
35 <sup>a</sup>	778	6.11	6.349933	-0.239933	6.092101	0.017899

<sup>a</sup> Test set compounds.

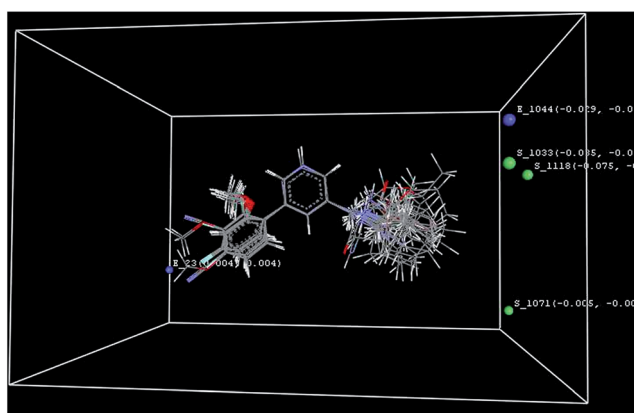
effect of electrostatic and steric interaction fields of the molecules on their NR2B binding affinity.

**Interpretation of 3D-QSAR models.** 3D QSAR studies generate 3D data points around molecules. The range of steric and electrostatic interaction field values in the generated 3D data points were used to interpret the 3D-QSAR models. The points generated in the *k*NN MFA SA 3D-QSAR model are E\_23 (0.0040 0.0040) S\_1033 (-0.0350 -0.0290) E\_1044 (-0.0290 -0.0190) S\_1071 (-0.0050 -0.0040) S\_1118 (-0.0750 -0.0550) and are shown in Fig. 9. The positive and negative range of the steric descriptor indicates that more steric substituents or groups and less steric substituents or groups are preferred for NR2B selectivity, respectively. Similarly, the positive and negative values of the electrostatic descriptors suggested the requirement of electropositive and electronegative groups, respectively, for NR2B selectivity. To interpret the 3D-QSAR model, the general structure of the pyrazine derivative was divided into three regions, such as (i) Central scaffold, (ii) Right hand side substitutions and (iii) Left hand side substitutions, as shown in Fig. 10.

**Central scaffold.** The planar aromatic central core is the minimum structural requirement of the molecules for NR2B selectivity. Various central scaffolds were studied by Brown *et al.*, 2011, *viz.* pyrazine, phenyl and three pyridine isomers.



(a) *k*NN-MFA SW-FB



(b) *k*NN-MFA SA

Fig. 9 Steric and electrostatic field points for SW-FB and SA methods.

Among the series of compounds, pyridine nitrogen isomer **18** ( $K_i$  12 nM) was found to be most potent, whereas another pyridine nitrogen isomer **21** ( $K_i$  8900 nM) was found to be least potent molecule. Brown *et al.* speculated that the reason for significant loss in activity of compound **21** was that the methoxy group and the pyridine lone pair create repulsive forces, which twist the methyl group out of a biologically relevant conformation.<sup>15</sup> Our computational results support this hypothesis; we studied the conformations of the most active compound **18** ( $K_i$  12 nM) and least active compound **21** ( $K_i$  8900 nM) by superimposing them on each other and found a significant difference in orientation of the methoxy group of each compound. Moreover, the aromatic ring bearing the methoxy group of compound **21** is flipped approximately 60 degrees relative to the aromatic ring of the actual bioactive conformation of the potent compound **18**, as shown in Fig. 11. In the case of pyridine, the lone pair does not contribute to the aromatic system. It points away from the ring, perpendicular to the 'p-orbital' in a region of localized electron density;<sup>52</sup> this may increase the steric bulk around the nitrogen of the pyridine ring. In compound **21**, this steric bulk around the pyridine nitrogen may come in close proximity to the methoxy group (lone pair present on oxygen atom). The steric requirements for the lone pair of electrons are higher than that for hydrogen, and hence repulsive forces may be generated between them. This could be one of the reasons for the change in orientation of the methoxy group and flipping of the phenyl ring. As the central scaffold requires planarity of the aromatic rings, perhaps it may be improved by incorporating more fused aromatic/hetero-aromatic rings at this position for optimization of binding affinity towards the NR2B subunit of the NMDA receptor.

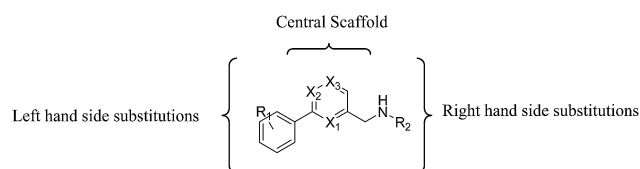


Fig. 10 Three regions of molecules considered in the study.

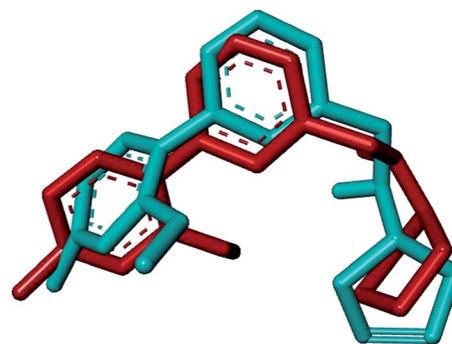


Fig. 11 Superimposition of compound **18** (red colour) and **21** (cyan colour). Hydrogen atoms of both structures are deleted for a better understanding.

**Right hand side substitutions.** 3D data points generated around this region are S\_1033 (−0.0350 −0.0290), S\_1071 (−0.0050 −0.0040), S\_1118 (−0.0750 −0.0550) for the steric interaction field and E\_1044 (−0.0290 −0.0190) for the electrostatic interaction field. The negative steric data points generated around this region indicate that less bulky groups, *i.e.*, nonaromatic rings (cyclic rings), are required at this position. The negative range of electrostatic points indicates that electronegative atoms, like oxygen, nitrogen, and sulphur, are favorable for NR2B selectivity at this position. A stepwise forward–backward method showed one electrostatic interaction field point (E\_1030 0.4580 0.4750) around this region. The positive range of this field value indicates that electropositive groups, like methyl, ethyl, propyl, isopropyl, cycloalkyl, and alkylamine, may be favorable at the right hand side of the central scaffold. Thus, the right hand side should accommodate substitutions that hold electronegative atoms with less steric electropositive groups.

**Left hand side substitutions.** This region was found to be surrounded by two 3D field electrostatic points E\_23 (0.0040 0.0040) for simulated annealing and another electrostatic point E\_202 (0.0380 0.0390) for the stepwise forward and backward

method. The positive range for electrostatic interaction at this position indicates that the presence of electropositive groups on the phenyl ring may contribute to defining NR2B selective NMDA antagonistic activity.

### 3.4 Designing of new compounds

The structural requirements derived from the 3D-QSAR studies were used to design novel active molecules (Table 8). Modifications were made on the central scaffold, right hand side chain and left hand side chain. It is noteworthy that the  $pK_i$  values of these designed compounds were found to be above 7. This indicates that the 3D-QSAR models may be a strategic guide to optimizing the poorly active molecules for use as NR2B selective NMDA receptor antagonists. To judge the reliability of the predicted activity of the designed compounds, all these compounds were docked into the NMDA receptor. Although we designed a series of compounds for use as NR2B selective NMDA antagonists, compounds with predicted NR2B binding ( $pK_i$ ) greater than 7 were considered to be active compounds and compounds with NR2B binding ( $pK_i$ ) less than 7 were defined as less active ones. Therefore, only six compounds (D1–D6) were selected for the docking study and rest were

Table 8 Structures of selected designed compounds and their predicted activities

Sr. no.	Compound	Predicted NR2B Binding $pK_i$	Physicochemical properties					Docking	
			log P	TPSA	MW	No. of H-bond acceptors	No. of H-bond donor	Volume	Binding energy kcal mol <sup>−1</sup>
D1		7.60	2.518	64.48	318.34	5	1	284.64	−9.6
D2		7.23	2.782	37.39	329.41	4	1	313.73	−8.8
D3		7.21	4.486	24.91	316.39	2	1	297.159	−10.4
D4		7.21	4.126	24.91	316.44	2	1	293.314	−9.2
D5		7.15	4.481	34.15	350.38	3	1	310.536	−10.2
D6		7.3	2.702	84.344	339.32	6	2	288.429	−10.2
	Comp 18 (most active in a series)	7.2	3.749	34.153	300.38	3	1	284.17	−9.3
	Ifenprodil (co-crystallized)	—	—	—	—	—	—	—	−11.3

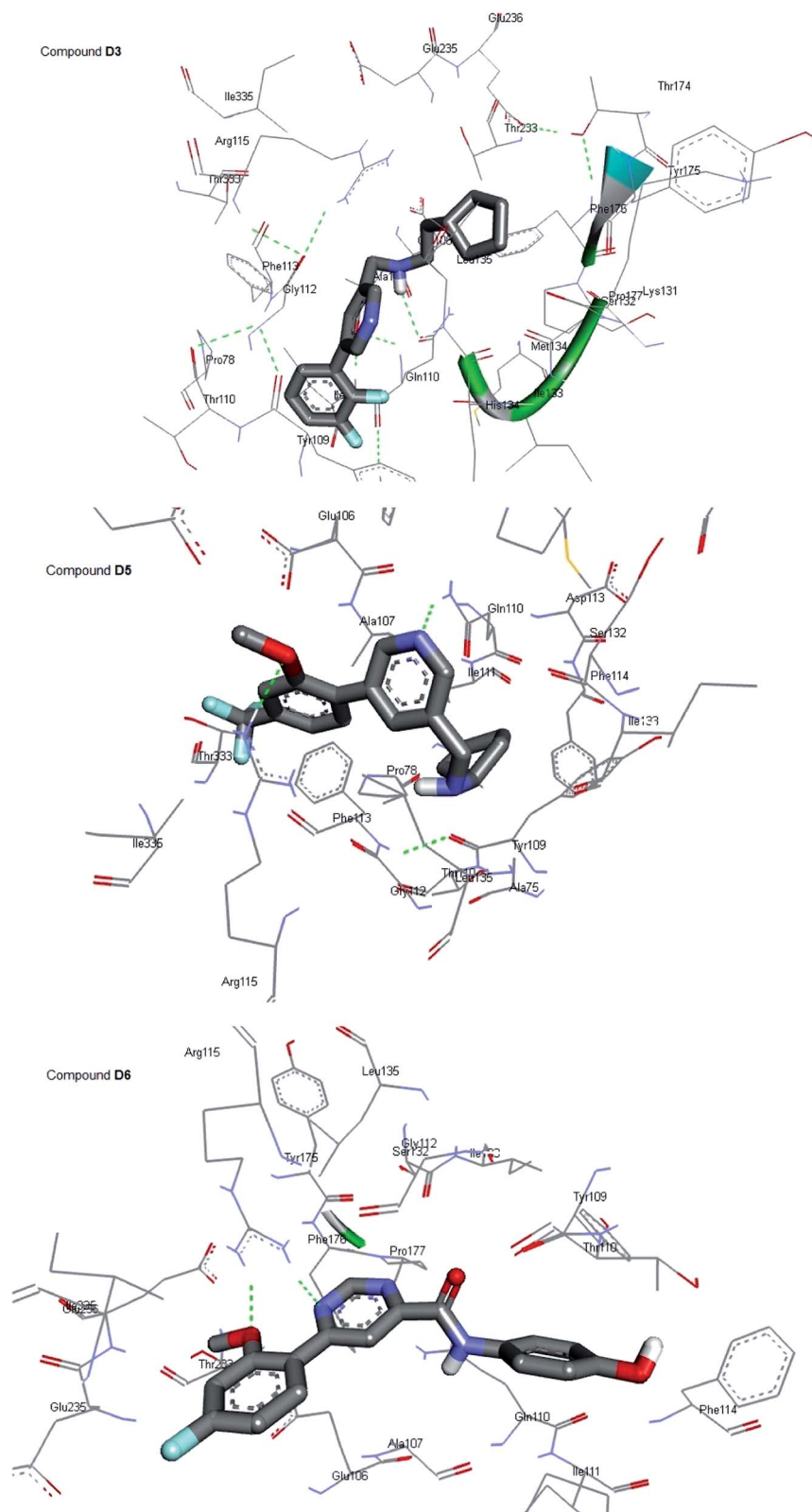


Fig. 12 Interacting models of D3, D5 and D6 with NMDA receptor.

excluded from it. It was found that the compounds suitably fit into the active site. The  $pK_i$  values predicted from 3D-QSAR models, as well as the binding energy calculated by the

AutoDock Vina docking program, are given in Table 8. The co-crystallized ligand (ifenprodil) and compound **18** are included in Table 8 as references to compare with the newly designed



Table 9 CYP450 Modulation by designed compounds

CYP450 Modulation					
Compound	CYP3A4	CYP2D6	CYP2C19	CYP2C9	CYP1A2
D1	Inhibitor (71% accuracy)	Inhibitor (81% accuracy)	Non-inhibitor (55% accuracy)	Non-inhibitor (71% accuracy)	Non-inhibitor (70% accuracy)
D2	Inhibitor (79% accuracy)	Inhibitor (91% accuracy)	Non-inhibitor (80% accuracy)	Non-inhibitor (86% accuracy)	Inhibitor (96% accuracy)
D3	Inhibitor (77% accuracy)	Inhibitor (86% accuracy)	Inhibitor (69% accuracy)	Non-inhibitor (76% accuracy)	Non-inhibitor (58% accuracy)
D4	Inhibitor (80% accuracy)	Inhibitor (93% accuracy)	Inhibitor (78% accuracy)	Non-inhibitor (56% accuracy)	Inhibitor (59% accuracy)
D5	Inhibitor (83% accuracy)	Inhibitor (93% accuracy)	Inhibitor (80% accuracy)	Non-inhibitor (81% accuracy)	Inhibitor (59% accuracy)
D6	Inhibitor (59% accuracy)	Non-inhibitor (67% accuracy)	Inhibitor (58% accuracy)	Inhibitor (76% accuracy)	Non-inhibitor (70% accuracy)

compounds. Although the 3D-QSAR predictions were not significantly strong, they still were comparable to reference compounds with respect to binding affinity calculated using docking studies. Compound **D3** and **D5** showed  $-10.4$  and  $-10.2$  kcal mol<sup>-1</sup> binding energy with the NMDA receptor, respectively. Interaction models for compound **D3** and **D5** are shown in Fig. 12. Compound **D3** establishes H-bond interaction between Gln110 and the -NH atom in the side chain of compound **D3** (Pose 3). Arg115 is also involved in H-bonding with the 'O' atom of the -OCH<sub>3</sub> group of compound **D5**, and there is also H-bond interaction between Gln110 and the central ring nitrogen atom. This indicates that Gln110 and Arg115 are key residues which take part in H-bonding to improve the binding affinity of ligands toward the NR2B subunit of the NMDA receptor. Compounds **D1**, **D2**, **D4** and **D6** also showed H-bonding interactions with Arg115. This could be the reason for their good binding affinities. Compound **D6** showed a binding energy of  $-10.2$  kcal mol<sup>-1</sup> and established H-bond interactions between the oxygen atom of the -OCH<sub>3</sub> group and Arg115 with a 1.904 Å distance. The pyridine nitrogen of **D6** also forms a H-bond with Arg115 with a 2.276 Å distance (Fig. 12). Although, compound **D6** is a completely different structure than ifenprodil, the shape acquired by **D6** was observed to be similar to that of ifenprodil. Therefore, shape-based features of this compound perhaps lead to improvement in p*K<sub>i</sub>* values and binding energy.

Physicochemical properties, such as log *P*, total polar surface area (TPSA), molecular weight (MW), H-bond acceptors, H-bond donor and molecular volume, were calculated for all the designed compounds<sup>53</sup> (Table 8). Secondary amine containing compounds may lead to variable pharmacokinetic profiles due to potent inhibition of the metabolizing enzyme cytochrome P450 (CYP) 2D6.<sup>15</sup> Compounds **D1–D5** possess secondary amines in their structures. Therefore, the CYP450 modulation was studied for these compounds using OCHEM<sup>54</sup> (Table 9). Compounds **D1–D5** were found to be inhibitors of CYP2D6. It was reported by Brown *et al.* that amide replacements of the secondary amine showed dramatic improvement in CYP2D6 inhibition. Hence, various compounds were designed by incorporating the amide group into the side-chain and studied

for their *in silico* CYP450 modulation. The designed compounds were found to be non-inhibitors of CYP2D6 but showed moderate NR2B binding p*K<sub>i</sub>* values (*i.e.* <7). To improve the NR2B binding p*K<sub>i</sub>* values, the compounds were further optimized by replacing the cyclic ring with an aromatic one. It was observed that compound **D6** showed a good NR2B binding p*K<sub>i</sub>* value (p*K<sub>i</sub>* 7.3) and retained non-inhibitor characteristics towards CYP2D6. Furthermore, compound **D6** was docked into the active site of NMDA receptor (PDB Id: 3QEL) by using AutoDock Vina and it was observed that compound **D6** fits well into the binding site with binding energy of  $-10.2$  kcal mol<sup>-1</sup>. This was considered a structure-based validation for the NR2B binding p*K<sub>i</sub>* value of compound **D6** predicted by the 3D-QSAR models.

## 4 Conclusions

The present work uncovers key structural requirements for NMDA inhibition employing various QSAR methods. The data set used in building 2D- and 3D-QSAR models comprises of structurally diverse NR2B selective NMDA receptor antagonists with wide biological activity variation. 2D-QSAR analyses have shown the major importance of Baumann's alignment independent topological descriptors in predicting NR2B selective NMDA receptor antagonistic activity of pyrazine and related derivatives. Out of two models (MLR and PLS), the MLR model was found to be the highly predictive, reliable model because of its significant external predictive power. Moreover, in the present study, 3D-QSAR models using *k*NN MFA SW-FB and SA methods were developed. Both 3D-QSAR models showed good internal as well as external predictive ability.

Elucidation of pharmacophoric features was carried out using a docking study. The putative pharmacophore model so obtained suggests that two aromatic rings, one H-bond donor and H-bond acceptor features of pyrazine derivatives would effectively contribute in the interaction mechanism with the NMDA receptor. 2D descriptors obtained from 2D-QSAR models, such as T\_C\_O\_7 and T\_T\_T\_6, and 3D data points obtained from 3D-QSAR show the importance of an electro-negative oxygen atom (*i.e.* acceptor) and phenyl ring (*i.e.*

aromatic ring). Hence, the descriptors from QSAR studies are found to corroborate with the pharmacophoric feature obtained from the docking studies. The pharmacophore model developed here could be used for virtual screening of commercially available databases, and the QSAR models derived from this study could be effectively used to predict and optimize the NMDA receptor binding activity of the lead compound obtained from the virtual screening. New compounds were designed so as to improve the NR2B subunit binding  $pK_i$  values using structural information obtained from 3D-QSAR models. The best predicted molecules were further docked into the active site of NMDA receptor using the AutoDock Vina program to observe their possible binding mode and interactions.

## Acknowledgements

The authors are thankful to Prof. M. N. Navale, Founder President of Sinhgad Technical Education Society, Pune for providing the necessary infrastructure. The authors gratefully acknowledge Dr Amit Nerkar for providing access to the VLife MDS software. We also thank Prof. R. B. Patil for his help in docking study.

## References

- 1 R. Gitto, L. D. Luca, S. Ferro, F. Occhiuto, S. Samper, G. D. Sarro, E. Russo, L. Ciranna, L. Costa and A. Chimirri, *ChemMedChem*, 2008, **3**, 1539–1548.
- 2 R. Gitto, L. D. Luca, S. C. Ferro, G. D. Sarro, L. Costa, L. Ciranna and A. Chimirri, *Bioorg. Med. Chem.*, 2009, **17**, 1640–1647.
- 3 J. Yosa, M. Blanco, O. Acevedo and L. R. Lareo, *Eur. J. Med. Chem.*, 2009, **44**, 2960–2966.
- 4 J. D. Davies, M. Crowe, N. Lucas, J. Quinn, D. D. Miller, S. Pritchard, D. Grose, E. Bettini, N. Calcinaghi, C. Virginio, L. Abberley, P. Goldsmith, A. D. Michel, I. P. Chessell, J. N. C. Kew, N. D. Miller and M. J. Gunthorpe, *Bioorg. Med. Chem. Lett.*, 2012, **22**, 2620–2623.
- 5 A. E. Geddes, X. F. Huang and K. A. Newell, *Prog. Neuro-Psychopharmacol. Biol. Psychiatry*, 2011, **35**, 896–904.
- 6 X. K. Wee, K. S. Ng, H. W. Leung, Y. P. Cheong, K. H. Kong, F. M. Ng, W. Soh, Y. Lam and C. M. Low, *Br. J. Pharmacol.*, 2010, **159**, 449–461.
- 7 M. E. Layton, M. J. Kelly and K. J. Rodzinak, *Curr. Top. Med. Chem.*, 2006, **6**, 697–709.
- 8 I. Borza, S. Kolok, A. Gere, J. Nagy, L. Fodor, K. Galgóczy, J. Fetter, F. Bertha, B. Agai, C. Horváth, S. Farkas and G. Domány, *Bioorg. Med. Chem. Lett.*, 2006, **16**, 4638–4640.
- 9 R. M. Santangelo, T. M. Acker, S. S. Zimmerman, B. M. Katzman, K. L. Strong, S. F. Traynelis and D. C. Liotta, *Expert Opin. Ther. Pat.*, 2012, **22**, 1337–1352.
- 10 M. Koller and S. Urwyler, *Expert Opin. Ther. Pat.*, 2010, **20**, 1683–1702.
- 11 C. Beinat, S. Banister, I. Moussa, A. J. Reynolds, C. S. McErlean and M. Kassiou, *Curr. Med. Chem.*, 2010, **17**, 4166–4190.
- 12 M. E. Layton, M. J. Kelly III, K. J. Rodzinak, P. E. Sanderson, S. D. Young, R. A. Bednar, A. G. Dilella, T. P. McDonald, H. Wang, S. D. Mosser, J. F. Fay, M. E. Cunningham, D. R. Reiss, C. Fandozzi, N. Trainor, A. Liang, E. V. Lis, G. R. Seabrook, M. O. Urban, J. Yergey and K. S. Koblan, *ACS Chem. Neurosci.*, 2011, **2**, 352–362.
- 13 I. Borza and G. Domány, *Curr. Top. Med. Chem.*, 2006, **6**, 687–695.
- 14 J. A. McCauley, C. R. Theberge, J. J. Romano, S. B. Billings, K. D. Anderson, D. A. Claremon, R. M. Freidinger, R. A. Bednar, S. D. Mosser, S. L. Gaul, T. M. Connolly, C. L. Condra, M. Xia, M. E. Cunningham, B. Bednar, G. L. Stump, J. J. Lynch, A. Macaulay, K. A. Wafford, K. S. Koblan and N. J. Liverton, *J. Med. Chem.*, 2004, **47**, 2089–2096.
- 15 D. G. Brown, D. L. Maier, M. A. Sylvester, T. N. Hoerter, E. Menhaji-Klotz, C. C. Lasota, L. T. Hirata, D. E. Wilkins, C. W. Scott, S. Trivedi, T. Chen, D. J. McCarthy, C. M. Maciag, E. J. Sutton, J. Cumberledge, D. Mathisen, J. Roberts, A. Gupta, F. Liu, C. S. Elmore, C. Alhambra, J. R. Krumrine, X. Wang, P. J. Ciaccio, M. W. Wood, J. B. Campbell, M. J. Johansson, J. Xia, X. Wen, J. Jiang, X. Wang, Z. Peng, T. Hu and J. Wang, *Bioorg. Med. Chem. Lett.*, 2011, **21**, 3399–3403.
- 16 B. Tewes, B. Frehland, D. Schepmann, K. U. Schmidtke, T. Winckler and B. Wünsch, *Bioorg. Med. Chem.*, 2010, **18**, 8005–8015.
- 17 B. Tewes, B. Frehland, D. Schepmann, K. U. Schmidtke, T. Winckler and B. Wünsch, *ChemMedChem*, 2010, **5**, 687–695.
- 18 C. J. McIntyre, J. A. McCauley, B. Bednar, R. A. Bednar, J. W. Butcher, D. A. Claremon, M. E. Cunningham, R. M. Freidinger, S. L. Gaul, C. F. Homnick, K. S. Koblan, S. D. Mosser, J. J. Romano and N. J. Liverton, *Bioorg. Med. Chem. Lett.*, 2009, **19**, 5132–5135.
- 19 I. Borza, E. Bozó, G. Barta-Szalai, C. Kiss, G. Tárkányi, A. Demeter, T. Gáti, V. Háda, S. Kolok, A. Gere, L. Fodor, J. Nagy, K. Galgóczy, I. Magdó, B. Agai, J. Fetter, F. Bertha, G. M. Keserü, C. Horváth, S. Farkas, I. Greiner and G. Domány, *J. Med. Chem.*, 2007, **50**, 901–914.
- 20 I. Borza, S. Kolok, G. Ignácz-Szendrei, I. Greiner, G. Tárkányi, K. Galgóczy, C. Horváth, S. Farkas and G. Domány, *Bioorg. Med. Chem. Lett.*, 2005, **15**, 5439–5441.
- 21 I. Borza, S. Kolok, A. Gere, E. Agai-Csongor, B. Agai, G. Tárkányi, C. Horváth, G. Barta-Szalai, E. Bozó, C. Kiss, A. Bielik, J. Nagy, S. Farkas and G. Domány, *Bioorg. Med. Chem. Lett.*, 2003, **13**, 3859–3861.
- 22 I. Borza, I. Greiner, S. Kolok, K. Galgóczy, G. Ignácz-Szendrei, C. Horváth, S. Farkas, T. Gáti, V. Háda and G. Domány, *Pharmazie*, 2006, **61**, 799–800.
- 23 M. Kawai, H. Nakamura, I. Sakurada, H. Shimokawa, H. Tanaka, M. Matsumizu, K. Ando, K. Hattori, A. Ohta, S. Nukui, A. Omura and M. Kawamura, *Bioorg. Med. Chem. Lett.*, 2007, **17**, 5533–5536.
- 24 M. Kawai, K. Ando, Y. Matsumoto, I. Sakurada, M. Hirota, H. Nakamura, A. Ohta, M. Sudo, K. Hattori, T. Takashima,

- M. Hizue, S. Watanabe, I. Fujita, M. Mizutani and M. Kawamura, *Bioorg. Med. Chem. Lett.*, 2007, **17**, 5558–5562.
- 25 E. Pinard, A. Alanine, A. Bourson, B. Buttellmann, M.-P. Heitz, V. Mutel, R. Gill, G. Trube and R. Wyler, *Bioorg. Med. Chem. Lett.*, 2002, **12**, 2615–2619.
- 26 C. F. Claiborne, J. A. McCauley, B. E. Libby, N. R. Curtis, H. J. Diggle, J. J. Kulagowski, S. R. Michelson, K. D. Anderson, D. A. Claremon, R. M. Freidinger, R. A. Bednar, S. D. Mosser, S. L. Gaul, T. M. Connolly, C. L. Condra, B. Bednar, G. L. Stump, J. J. Lynch, A. Macaulay, K. A. Wafford, K. S. Koblan and N. J. Liverton, *Bioorg. Med. Chem. Lett.*, 2003, **13**, 697–700.
- 27 K. T. Nguyen, C. F. Claiborne, J. A. McCauley, B. E. Libby, D. A. Claremon, R. A. Bednar, S. D. Mosser, S. L. Gaul, T. M. Connolly, C. L. Condra, B. Bednar, G. L. Stump, J. J. Lynch, K. S. Koblan and N. J. Liverton, *Bioorg. Med. Chem. Lett.*, 2007, **17**, 3997–4000.
- 28 N. J. Liverton, R. A. Bednar, B. Bednar, J. W. Butcher, C. F. Claiborne, D. A. Claremon, M. Cunningham, A. G. DiLella, S. L. Gaul, B. E. Libby, E. A. Lyle, J. J. Lynch, J. A. McCauley, S. D. Mosser, K. T. Nguyen, G. L. Stump, H. Sun, H. Wang, J. Yergey and K. S. Koblan, *J. Med. Chem.*, 2007, **50**, 807–819.
- 29 V. P. Zambre, P. R. Murumkar, R. Giridhar and M. R. Yadav, *J. Chem. Inf. Model.*, 2009, **49**, 1298–1311.
- 30 V. P. Zambre, P. R. Murumkar, R. Giridhar and M. R. Yadav, *J. Mol. Graphics Modell.*, 2010, **29**, 229–239.
- 31 V. P. Zambre, R. Giridhar and M. R. Yadav, *Med. Chem. Res.*, 2013, **22**, 4685–4699.
- 32 N. P. Dixit, P. Devre, S. D. Sawant and V. P. Zambre, *Int. J. Pharm. Pharm. Sci.*, 2013, **5**, 569–573.
- 33 P. R. Murumkar, S. DasGupta, V. P. Zambre, R. Giridhar and M. R. Yadav, *Chem. Biol. Drug Des.*, 2009, **73**, 97–107.
- 34 P. R. Murumkar, V. P. Zambre and M. R. Yadav, *J. Comput.-Aided Mol. Des.*, 2010, **24**, 143–156.
- 35 A. M. Kanhed, V. P. Zambre, V. A. Pawar, M. K. Sharma, R. Giridhar and M. R. Yadav, *Med. Chem. Res.*, 2014, **23**, 5215–5223.
- 36 O. Trott and A. J. Olson, *J. Comput. Chem.*, 2010, **31**, 455–461.
- 37 *VLife MDS, version 4.3*, VLife Sciences Technologies Pvt. Ltd., Pune, India, 2008.
- 38 J. Gasteiger and M. Marsili, *Tetrahedron*, 1980, **36**, 3219–3228.
- 39 P. Prathipati and A. K. Saxena, *SAR QSAR Environ. Res.*, 2003, **14**, 433–446.
- 40 C. Croux and K. Joossens, *J. Multivariate Anal.*, 2005, **96**, 348–403.
- 41 J. Devillers, *Neuronal network in QSAR and drug design*, Academic Press, London, 1996.
- 42 D. Ivan, L. Crisan, S. Funar-Timofei and M. Mracec, *J. Serb. Chem. Soc.*, 2012, **77**, 1–25.
- 43 H. J. Smith and H. Williams, *Introduction to the Principles of Drug Design and Action*, 4th edn, 2005, pp. 185–210.
- 44 S. Wold, A. Ruhe, H. Wold and W. J. Dunn, *SIAM J. Sci. Stat. Comp.*, 1984, **5**, 735–743.
- 45 M. Shen, A. LeTiran, Y. Xiao, A. Golbraikh, H. Kohn and A. Tropsha, *J. Med. Chem.*, 2002, **4**, 2811–2823.
- 46 A. Golbraikh and A. Tropsha, *J. Mol. Graphics Modell.*, 2002, **20**, 269–276.
- 47 M. Shen, Y. Xiao, A. Golbraikh, V. K. Gombar and A. Tropsha, *J. Med. Chem.*, 2003, **46**, 3013–3020.
- 48 A. Golbraikh and A. Tropsha, *J. Comput.-Aided Mol. Des.*, 2002, **16**, 357–369.
- 49 P. P. Roy and K. Roy, *QSAR Comb. Sci.*, 2008, **27**, 302–313.
- 50 P. P. Roy, S. Paul, I. Mitra and K. Roy, *Molecules*, 2009, **14**, 1660–1701.
- 51 J. Devillers and R. Lipinick, Practical Applications of Regression Analysis in Environmental QSAR Studies, in *Practical Applications of Quantitative Structure Activity Relationships (QSAR). Practical Applications of Quantitative Structure Activity Relationships (QSAR) in Environmental chemistry and Toxicology*, ed. W. Kracher and J. Devillers, Kluwer, Dordrecht, 1990, pp. 129–143.
- 52 S. A. Lawrence, Heterocyclic Amines, in *Amines: Synthesis, Properties and Applications*, Cambridge University Press, 2004, p. 120.
- 53 <http://www.molinspiration.com>.
- 54 I. Sushko, S. Novotarskyi, R. Körner, A. K. Pandey, M. Rupp, W. Teetz, S. Brandmaier, A. Abdelaziz, V. V. Prokopenko, V. Y. Tanchuk, R. Todeschini, A. Varnek, G. Marcou, P. Ertl, V. Potemkin, M. Grishina, J. Gasteiger, C. Schwab, I. I. Baskin, V. A. Palyulin, E. V. Radchenko, W. J. Welsh, V. Kholodovych, D. Chekmarev, A. Cherkasov, J. Aires-de-Sousa, Q. Y. Zhang, A. Bender, F. Nigsch, L. Patiny, A. Williams, V. Tkachenko and I. V. Tetko, *J. Comput.-Aided Mol. Des.*, 2011, **25**, 533–554.